SPRING FORECASTING EXPERIMENT 2020

Conducted by the

EXPERIMENTAL FORECAST PROGRAM

of the

NOAA/HAZARDOUS WEATHER TESTBED

HWT Facility – National Weather Center

27 April - 29 May 2020

https://hwt.nssl.noaa.gov/sfe/2020/

Program Overview and Operations Plan

2 May, 2020

Adam Clark[2], Israel Jirak[1], Burkely T. Gallo[1,3], Andy Dean[1], Kent Knopfmeier[2,3], Brett Roberts[1,2,3], Louis Wicker[2], Makenzie Krocak[3,4], Patrick Skinner[2,3], Pam Heinselman[2], Katie Wilson[2,3], Jake Vancil[1,3], Kimberly Hoogewind[2,3], Nathan Dahl[1,3], Gerry Creager[2,3], Thomas Jones[2,3], Jidong Gao[1], Yunheng Wang[2,3], Eric D. Loken[2,3,4], Montgomery Flora[2,3,4], Chris Kerr[2,3], Nusrat Yussouf[2,3], Scott Dembek[2,3], William Miller[2,3], Joshua Martin[2,3], Jorge Guerra[2,3], Brian Matilla[2,3], David Jahn[1,3], David Harrison[1,3], and Dave Imy[2]

(1) NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma
(2) NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma
(3) Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma
(4) School of Meteorology, University of Oklahoma, Norman, Oklahoma

Table of Contents

The NOAA Hazardous Weather Testbed (photo credit: James Murnan, NSSL)

1. Introduction

Each spring, the Experimental Forecast Program (EFP) of the NOAA/Hazardous Weather Testbed (HWT), organized by the Storm Prediction Center (SPC) and National Severe Storms Laboratory (NSSL), conducts a collaborative experiment to test emerging concepts and technologies designed to improve the prediction of hazardous convective weather. The primary goals of the HWT are to accelerate the transfer of promising new tools from research to operations, to inspire new initiatives for operationally relevant research, and to identify and document sensitivities and the performance of state-of-the art experimental convection-allowing (1 to 3 km grid-spacing) modeling systems.

The 2020 Spring Forecasting Experiment (SFE 2020), a cornerstone of the EFP, will be conducted 27 April – 29 May. Because of the COVID-19 pandemic, restrictions on travel and gatherings preclude an in-person experiment in the HWT. However, to maintain momentum in several areas of convection-allowing model (CAM) development, the EFP will perform a limited, virtual experiment with external participants focused on next-day evaluations of model performance. In addition, local staff and National Weather Service (NWS) forecasters will perform a small number of online real-time forecasting activities in the afternoon focused on Warn-on-Forecast applications. The pandemic will not impact contributions of experimental CAM datasets that were planned for SFE 2020. Thus, these systems will run as originally planned, and for the 5^{th} year have been coordinated into a single ensemble framework called the Community Leveraged Unified Ensemble (CLUE; Clark et al. 2018). The 2020 CLUE is constructed by using a set of common model specifications (e.g., grid-spacing, domain size, model version, etc.) so that the simulations contributed by each group can be used in carefully designed controlled experiments. This design will once again allow us to conduct several experiments geared toward identifying optimal configuration strategies for CAM ensembles. The 2020 CLUE includes 41 members with grid-spacing $\leq$ 3-km. SFE 2020 will also involve the continued testing of a 3-km grid-spacing Warn-on-Forecast prototype system (WoFS Ensemble, hereafter), which will be used for the fourth year to issue very short lead-time outlooks. Additional WoFS Ensemble simulations will run with 1.5-km grid-spacing to test the impact of more refined grid-spacing.

This operations plan summarizes the core interests of SFE 2020 with information on experiment operations. The organizational structure of the HWT and information on various forecast tools and diagnostics can also be found in this document. The remainder of the operations plan is organized as follows: Section 2 provides details on model and products being tested during SFE 2020 and Section 3 describes the core interests and new concepts being introduced for SFE 2020. A list of daily participants, details on the SFE forecasting, and more general information on the HWT are found in appendices.

2. Overview of Experimental Products and Models

Daily model evaluation activities will occur from 10am – noon (CDT) focusing on various CLUE subsets. The 2020 CLUE includes recent versions of the Advanced Research Weather Research and Forecasting (WRF-ARW) model, deterministic and ensemble forecasts based on the United Kingdom Met Office's Unified Modeling System, and several stand-alone-regional configurations of the Finite Volume Cubed-Sphere model (FV3-SAR). In addition to the CLUE, the operational 3-km grid-spacing High-Resolution Ensemble Forecast system Version 2.1 (HREFv2.1) will be examined, along with its updated version, HREFv3. The rest of this section provides further details on each modeling system utilized in SFE 2020.

a) The 2020 Community Leveraged Unified Ensemble (CLUE)

The CLUE is a carefully designed ensemble with members contributed by NSSL, NOAA's Environmental Modeling Center (EMC), NOAA's Global Systems Laboratory (GSL), NCAR, NOAA's Geophysical Fluid Dynamics Laboratory (GFDL), and the United Kingdom Meteorology Office (UK Met). Most members are initialized weekdays at 0000 UTC, while members within experiments testing different time-lagging strategies are initialized at various other times. CLUE members have 3-km grid-spacing and CONUS domain, except for the UK Met members which use 2.2 km grid-spacing and slightly sub-CONUS domain. Depending on the CLUE subset, forecast lengths range from 12 to 60 h. Table 1 summarizes all 2020 CLUE contributions. Subsequent tables provide details on members in each subset.

Table 1 Summary of the 12 unique subsets that comprise the 2020 CLUE.

Clue Subset	**# of mems**	**IC/LBC perts**	**Mixed Physics**	**Data Assimilation**	**Model Core**	**Agency**	**Init. Time(s) UTC**
HRRRv4	1	none	no	GSI-EnVar	ARW	GSL	00-23
HRRRE	9	EnKF	no	EnKF	ARW	GSL	00, 06, 12, 18
gsl-fv3sar	4	none	yes	cold start	FV3	GSL	00
arw-ICs	2	none	no	cold start	ARW	NCAR	00
ukmet	9	MOGREPS-G	no	cold start	UM	UK Met	18, 00
um-ICs	2	none	no	cold start	UM	UK Met	00
nssl-glm	1	none	no	NSSL-VAR (GLM)	ARW	NSSL	00
nssl-noglm	1	none	no	NSSL-VAR (no GLM)	ARW	NSSL	00
nssl-tl	6	none	yes	cold start	ARW	NSSL	02, 03, 05, 08, 09, 11
sarfv3-ICs	2	none	no	cold start	FV3	NSSL	00
emc-fv3sar	3	none	no	cold start	FV3	EMC	00
gfdl-fv3	1	none	no	cold start	FV3	GFDL	00

Table 2 Specifications for the hrrrv4 CLUE member. The next and final update to the deterministic Rapid Refresh, version 5 (RAPv5), and HRRRv4, is scheduled for an operational implementation in June 2020. The physics suite for HRRRv4 continues to use actively-developed versions of Thompson et al. (2014) aerosol-aware microphysics, MYNN PBL scheme, RUC land surface model and RRTMG SW/LW radiation schemes. Enhancements have been made to the MYNN PBL scheme to further improve both representation of sub-grid-scale clouds and their effects on the local environment (reducing model bias of incoming radiation and temperature/moisture fields). Gravity-wave drag enhancements have been made to improve representation of the effects of sub-grid terrain on the horizontal flow. Land surface model and state changes include installation of an inland lake model for improved lake-temperature prediction, higher-resolution MODIS albedo and inland lake datasets, use of fractional sea-ice data and FVCOM dynamic specification of temperature and ice concentrations for the Great Lakes. Finally, VIIRS-based fire-radiative power detections are used to specify wildfire-driven injection of particulate matter for 3-D advection and deposition of smoke plumes. Enhancements to numerics in HRRRv4 include a reduction in magnitude of the 6th order filter for momentum, thermodynamic and hydrometeor fields to improve depiction of weaker small-scale cloud and precipitation features. A new implicit-explicit vertical advection scheme in HRRRv4 permits larger vertical motion in intense convection to facilitate improved diagnosis of rotational features such as mesocyclones. For data assimilation, The HRRRv4 uses an updated version of GSI and includes assimilation of additional datasets including lightning data from GOES (GLM), aircraft and RAOB moisture observations above 300 mb, and tropical cyclone central pressure estimates from TCvitals for improved position and structure of tropical systems. A 36-member, hourly-cycled, storm-scale ensemble data assimilation system (HRRRDAS) provides a background deterministic state estimate (ensemble mean) and background ensemble for initialization of the CONUS HRRRv4. This system is designed to improve use of conventional and radar observations during data assimilation with better representation of meso-to-storm scale covariances when compared with the comparatively coarse global ensemble (GDAS) used in HRRRv3. More accurate retention and evolution of meso-to-storm scale features, particularly in the early forecast hours, are intended benefits of HRRRDAS use. The HRRRDAS, while intended to improve deterministic HRRRv4 forecasts, also forms the basis for HRRR ensemble forecasts described in the HRRRE section.

Member: HRRRv4	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
hrrrv4	HRRRDAS	RAP	Thompson	MYNN	RUC	RRTMG	ARW

Table 3 Specifications for the HRRRE CLUE members. During spring 2020, GSL will run a parallel version of the 36-member, hourly-cycled, 3-km ensemble ("HRRR Data-Assimilation System" or "HRRRDAS") that is a component of the HRRRv4 operational implementation. HRRRDAS members are initialized at 0900 and 2100 UTC from a combination of RAP for atmospheric ensemble mean, GDAS for atmospheric perturbations, and RAP/HRRR for land surface. The 36 HRRRDAS members are cycled hourly with GSI-EnKF to assimilate conventional and radar-reflectivity observations. The hourly cycling also includes cloud clearing. Experimental HRRR Ensemble (HRRRE) forecasts are initialized from the first 9 HRRRDAS members every 6 h. Posterior inflation during the hourly cycling, random boundary-condition perturbations, and stochastic parameter perturbations (SPP) applied to the land-surface, PBL, and microphysics schemes contribute to spread of the HRRRE forecasts. The HRRRDAS analyses and HRRRE forecasts provide initial conditions and boundary conditions for the experimental Warn-on-Forecast system. Additional HRRRDAS/HRRRE documentation is available at https://rapidrefresh.noaa.gov/internal/pdfs/2020_Spring_Experiment_HRRRE_Documentation.pdf.

Members: HRRRE	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
hrrre01	enkf_m01b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre02	enkf_m02b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre03	enkf_m03b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre04	enkf_m04b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre05	enkf_m05b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre06	enkf_m06b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre07	enkf_m07b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre08	enkf_m08b	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre09	enkf_m09b	GFS	Thompson	MYNN	RUC	RRTMG	ARW

Table 4 Specifications for the gsl-sarfv3 CLUE members. These runs will use the stand-alone-regional (SAR) FV3 code contained within the UFS code base maintained by EMC as it was 4/9/2020. Each run uses the HRRRv4 physics package. The "Hord" column refers to the horizontal advection method. Hord = 5 is the fastest and least diffusive option, while Hord = 6 is a developmental scheme with an intermediate strength monotonicity constraint and is more diffusive than Hord = 5.

Member: gsl-sarfv3	ICs	LBCs	Micro-physics	PBL	LSM	Radiation	Hord	Model
gsl-fv3sar01	HRRRv4	RAP	Thompson	MYNN	RUC	RRTMG	5	FV3
gsl-fv3sar02	HRRRv4	RAP	Thompson	MYNN	RUC	RRTMG	6	FV3
gsl-fv3sar03	GFS	GFS	Thompson	MYNN	RUC	RRTMG	5	FV3
gsl-fv3sar04	GFS	GFS	Thompson	MYNN	RUC	RRTMG	6	FV3

Table 5 Specifications for the arw-ICs CLUE members. NCAR is providing 36-h forecasts initialized at 0000 UTC from GFS and UM analyses using version 3.6.1 of the WRF-ARW model over a 3-km domain spanning the CONUS. Both forecasts have 40 vertical levels with a 50-hPa top and use Thompson microphysics, MYJ PBL, Noah land-surface model, and RRTMG shortwave and longwave radiation with aerosol and ozone climatologies. In short, the two forecasts are identical except for their ICs and LBCs.

Members: arw-ICs	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Model
arw-ICs01	GFS	GFSf	Thompson	MYJ	NOAH	RRTMG	ARW
arw-ICs02	UM	UMf	Thompson	MYJ	NOAH	RRTMG	ARW

Table 6 Specifications for the ukmet CLUE members. These single physics members, which use the Met Office Unified Model (UM) and Joint UK Land Environment Simulator (JULES), have 2.2 km grid-spacing with 70 vertical levels covering a slightly sub-CONUS domain. The members are closely aligned with the UK ensemble, MOGREPS-UK. The members are cold-start initialized from 3-h forecasts of 1800 and 0000 UTC initializations of the Met Office global ensemble system, MOGREPS-G, and provide 48-h forecasts. Model uncertainty is depicted by the Random Parameter (RP) scheme, which stochastically perturbs a subset of physics parameters from the PBL and microphysics schemes throughout the forecast. The PBL scheme consists of a 3D turbulent mixing scheme using a locally scale-dependent blending of Smagorinsky and non-local K-profile boundary layer mixing schemes, and the Smith cloud scheme is used, where partial cloudiness is diagnosed assuming a trianglular moisture distribution with a width that is a universally specified function of height only. There is no convection-parameterization and single-moment microphysics is used. The "RA config" column refers to the "Regional Atmosphere" configuration. For these members "RA2M" is used, which is designed for mid-latitudes.

Members: ukmet	ICs/LBCs	RA config	Micro-physics	PBL	Cloud Scheme	Model
ukmet01	MOGREPS-G01	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet02	MOGREPS-G02	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet03	MOGREPS-G03	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet04	MOGREPS-G04	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet05	MOGREPS-G05	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet06	MOGREPS-G06	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet07	MOGREPS-G07	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet08	MOGREPS-G08	RA2M	single-mom.	Smag. blended	Smith	UM
ukmet09	MOGREPS-G09	RA2M	single-mom.	Smag. blended	Smith	UM

Table 7 Specifications for the um-ICs CLUE members contributed by the UK Met Office. This member uses the same configuration as the ukmet members listed in Table 6, with the only difference being that ICs/LBCs are from the GFS and deterministic version of UK Met's global model.

Members: um-ICs	ICs/LBCs	RA config	Micro-physics	PBL	Cloud Scheme	Model
um-ICs01	GFS	RA2M	single-mom.	Smag. blended	Smith	UM
um-ICs02	UM deterministic	RA2M	single-mom.	Smag. blended	Smith	UM

Table 8 Specifications for the nssl-glm and nssl-noglm CLUE members contributed by NSSL. These runs use the HRRRv4 physics and RAPv3 as the background for a data assimilation system known as NEWS3DVAR (Fierro et al. 2016, 2018) that assimilates radar data and is run with and without total lightning data from the GOES Geostationary Lightning Mapper (GLM). These forecasts are run to 12-h over a CONUS domain with 3-km grid-spacing.

Members: nssl-glm	ICs	LBCs	Micro-physics	PBL	LSM	Radiation	Model
nssl-glm	NEWS3DVAR	RAP	Thompson	MYNN	RUC	RRTMG	ARW
nssl-noglm	NEWS3DVAR	RAP	Thompson	MYNN	RUC	RRTMG	ARW

Table 9 Specifications for the nssl-tl members contributed by NSSL. These members are initialized from the HRRRv4 ICs at 02, 03, 05, 08, 09, and 11 UTC as part of a time-lagged ensemble. All members use 3 km grid-spacing. Members nssl-tl02 and nssl-tl05 use the HRRRv4 physics package, and the other members use the NSSL-WRF physics configuration..

Members: nssl-tl	ICs	LBCs	Micro-physics	PBL	LSM	Radiation	Forecast length (h)	Model
nssl-tl01	02Z HRRRv4	GFSf	WSM6	MYJ	NOAH	RRTM/Dudhia	34	ARW
nssl-tl02	03Z HRRRv4	GFSf	Thompson	MYNN	RUC	RRTMG	33	ARW
nssl-tl03	05Z HRRRv4	GFSf	WSM6	MYJ	NOAH	RRTM/Dudhia	31	ARW
nssl-tl04	08Z HRRRv4	GFSf	WSM6	MYJ	NOAH	RRTM/Dudhia	28	ARW
nssl-tl05	09Z HRRRv4	GFSf	Thompson	MYNN	RUC	RRTMG	27	ARW
nssl-tl06	11Z HRRRv4	GFSf	WSM6	MYJ	NOAH	RRTM/Dudhia	25	ARW

Table 10 Specifications for the sarfv3-ICs members contributed by NSSL. Both sarfv3-ICs members will use the SAR FV3 code contained within the UFS code base maintained by EMC as it was 4/9/2020. Both runs use an experimental configuration of 80 vertical levels.

Members: nssl-glm	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Model
sarfv3-ICs01	GFS	GFSf	Thompson	MYNN	NOAH	RRTMG	FV3
sarfv3-ICs02	UM	UMf	Thompson	MYNN	NOAH	RRTMG	FV3

Table 11 Specifications for the emc-fv3sar CLUE members. All three members use the SAR FV3 code contained within the UFS code base maintained by EMC as it was 4/9/2020. For SFE 2020, EMC plans to provide three configurations of this system at 3-km grid-spacing and 50 vertical levels. The first is the control version of the 3 km limited area, SAR configuration which has been running at EMC for about 2 years and utilizes the GFSv15 physics suite (emc-fv3sar). The second is an experimental version of the SAR configuration which will use the Thompson microphysics and MYNN planetary boundary layer schemes (emc-fv3sarX). The third configuration is similar to the second configuration but includes an hourly data assimilation cycle over a 6-hour period prior to issuance of a free forecast (emc-fv3sarDA). EMC will leverage these three configurations to evaluate the impacts of running in a limited area capacity. The output grid is identical to that used by the NAM CONUSnest and HRRR domains. Each run is initialized daily at 0000 UTC with forecasts to 60 h using initial and 3-hourly lateral boundary conditions from the 0000 UTC GFSv15 system currently in operations.

Members: emc-fv3sar	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
emc-fv3sar	GFSv15	GFSv15f	GFDL	EDMF	NOAH	RRTMG	FV3
emc-fv3sarX	GFSv15	GFSv15f	Thompson	MYNN	NOAH	RRTMG	FV3
emc-fv3sarDA	GFSv15	GFSv15f	Thompson	MYNN	NOAH	RRTMG	FV3

Table 12 Specifications for the gfdl-fv3 CLUE member. The GFDL configuration uses a combination of grid nesting (Harris and Lin, 2013) and stretching (Harris et al 2016) to refine a 13-km global grid to a 3-km nested grid covering the CONUS region. This model consists of FV3 coupled to a modified form of the GFS Physics (Chen et al 2019 and references therein) and the Noah land model. Yonsei University PBL scheme (Hong et al 2006) and the six-category single-moment GFDL microphysics (Zhou et al 2019) are used. The deep convective scheme is disabled on the nested grid. Initialization is a cold-start from regridded GFS real-time analyses. GFDL will provide simulations run daily at 00Z out to 126 hours to demonstrate the potential for medium-range prediction of convective-scale events. GFDL has made a variety of updates to the model since 2019 including continued revisions to the GFDL microphysics and associated cloud-radiation interactions; revisions to YSU to better handle higher vertical resolutions; and improvements to the advection schemes and two-way nesting.

Member: gfdl-fv3	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
gfdl-fv3	GFS	n/a	GFDL	YSU	NOAH	RRTMG	FV3

The configuration of the 2020 CLUE will allow for several unique experiments that have been designed to examine issues immediately relevant to the design of a NCEP/EMC operational CAM-based ensemble. Some of these experiments are listed below:

Model initial condition vs. model core sensitivity: This experiment involves multi-agency, international collaboration among NCAR, NSSL, and the UK Met Office. A matrix of 6 simulations will be examined, with NCAR, NSSL, and UK Met running WRF-ARW (Table 5), SAR-FV3 (Table 10), and UM (Table 7), respectively, initialized from both the GFS and UM global models. The goal of this comparison is to examine forecast sensitivity to different initial conditions and model cores at convective scales.

FV3-SAR Configurations: GSL, NSSL, and EMC will run various configurations of the FV3-SAR (Tables 4, 10, and 11, respectively). These coordinated runs will allow for the assessment of many aspects of FV3-SAR configuration: physics, data assimilation, number of vertical levels, initial conditions, and horizontal advection settings.

Single-model time-lagging: This experiment is designed to test single-model ensemble time-lagging strategies relative to an ensemble composed of members initialized from a single time. The 1200 UTC

HRRRE (Table 3) will be compared to a HRRRE time-lagged ensemble, and a time-lagged ensemble based on deterministic HRRRv4 (Table 2) and NSSL-WRF (Table 9) configurations.

Multi-model vs. time-lagging: This experiment will involve several sets of comparisons based at 0000 UTC in which HREF will be used as a baseline to assess (1) single-model ensembles initialized from one time, (2) single model ensembles that are time-lagged, and (3) multi-model ensembles that are time-lagged. Each of the experiments will involve various combinations of 1800 and 0000 UTC initializations of HRRRE (Table 3) and the UK Met UM (Table 6).

Total Lightning Data Assimilation: The goal of this experiment is to assess the value of assimilating total lightning data from the GOES-16 Geostationary Lightning Mapper (GLM) in regions with poor radar coverage by comparing two 12-h WRF runs: one with assimilation of GLM data and one without (Table 8).

To ensure consistent post-processing, visualization, and verification for subsets of CLUE ensemble members contributed by different collaborators, all groups will generally utilize similar post-processing software to output the same set of model output fields on the same grid. For WRF-ARW and FV3-SAR members, the Unified Post-Processor software (UPP; available at http://www.dtcenter.org/upp/users/downloads/index.php) is used to output a minimum set of 123 output fields from each CLUE member. These fields (output in grib2 format) are the same as the 2D fields output by HRRRv3 and were chosen because of their relevance to a broad range of forecasting needs, including aviation, severe weather, and precipitation. The UM ensembles will output a much smaller set of fields limited to low-level temperature, dewpoint, and winds; lowest model level and composite reflectivity, hourly maximum 2-5 km AGL updraft helicity, and total precipitation. A table listing the output diagnostics for ARW is included in Appendix C.

b) High Resolution Ensemble Forecast (HREFv2.1) System

HREFv2.1 is a 10-member CAM ensemble currently running at EMC (forecasts viewable at: http://www.spc.noaa.gov/exper/href/). HREFv2 was implemented operationally on 1 November 2017 and was recently updated to include two HRRR members (one 6-h time lagged). The design of HREFv2.1 originated from the SSEO, which demonstrated skill for six years in the HWT and SPC prior to HREFv2 operational implementation. All members, except for the NAM CONUS Nest and HRRR, are initialized with a “cold-start”. Forecasts to 36 h are produced at 0000 and 1200 UTC. HREFv2.1 specifications are listed below in Table 13.

Table 13 Model specifications for HREFv2.1.

HREFv2.1	**ICs**	**LBCs**	**Micro-physics**	**PBL**	**dx (km)**	**Vertical Levels**	**Included in HREF hours**
HRRRv3	RAP -1h	RAP -1h	Thompson	MYNN	3.0	50	0 – 36
HRRRv3 -6h	RAP -1h	RAP -1h	Thompson	MYNN	3.0	50	0 – 30
HRW ARW	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 48
HRW ARW -12h	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 36
HRW NMMB	RAP	GFS -6h	Ferrier-Aligo	MYJ	3.2	50	0 – 48
HRW NMMB -12h	RAP	GFS-6h	Ferrier-Aligo	MYJ	3.2	50	0 – 36
HRW NSSL	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 48
HRW NSSL -12h	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 36
NAM CONUS Nest	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 48
NAM CONUS Nest -12h	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 48

c) High Resolution Ensemble Forecast (HREFv3) System

HREFv3 replaces the HRW NMMB simulations with *emc-fv3sar* (Table 11) and HRRRv3 with HRRRv4.

Table 14 Model specifications for HREFv3.

HREFv3	**ICs**	**LBCs**	**Micro-physics**	**PBL**	**dx (km)**	**Vertical Levels**	**Included in HREF hours**
HRRRv4	HRRRDAS	RAP -1h	Thompson	MYNN	3.0	50	0 – 36
HRRRv4 -6h	HRRRDAS	RAP -1h	Thompson	MYNN	3.0	50	0 – 30
HRW ARW	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 48
HRW ARW -12h	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 36
HRW FV3	GFS	GFS -6h	GFDL	EDMF	3	50	0 – 60
HRW FV3 -12h	GFS	GFS-6h	GFDL	EDMF	3	50	0 – 48
HRW NSSL	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 48
HRW NSSL -12h	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 36
NAM CONUS Nest	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 60
NAM CONUS Nest -12h	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 48

d) NSSL Warn-on-Forecast Experiments

The NSSL Experimental Warn-on-Forecast System (WoFS) is a 36-member WRF-based ensemble data assimilation system used to produce very short-range (0-6 h) probabilistic 18-member forecasts of hazardous weather phenomena such as supercell thunderstorm rotation, hail, high winds, and flash flooding. The 900-km wide daily WoFS domain will target the primary region where severe weather is anticipated. The 2020 SFE will utilize the following three WoFS system configurations, described in detail below: (i) Real-time; (ii) High-resolution; and (iii) Dual Resolution Hybrid DA.

The starting point for each day's experiment will be the experimental High-Resolution Rapid Refresh Data Assimilation System (HRRRDAS) and the HRRRE (Table 3) provided by GSL. A 1-h forecast from the 1400 UTC, 36-member, hourly-cycled HRRRDAS analysis provides the initial conditions for all three WoFS configurations. Boundary conditions for the WoFS are provided by the combination of the 0600 UTC and 1200 UTC HRRRE forecasts, which are initialized from the HRRRDAS analyses at these times, for the period 1500 UTC Day 1 – 0300 UTC Day 2.

i) Real-time WoFS

The full 36-member real-time WoFS ensemble, run from 1500 UTC Day 1 to 0300 UTC Day 2, will be updated every 15 minutes by GSI-EnKF data assimilation of MRMS radar reflectivity and radial velocity data, cloud water path retrievals, atmospheric motion vectors, and clear-sky radiances from the GOES-16 imager, data from the GOES-16 Geostationary Lightning Mapper (GLM), and Oklahoma Mesonet observations (when available). Conventional (i.e. prepbufr) observations will also be assimilated at 15 minutes past each hour. All real-time WoFS ensemble members utilize the NSSL 2-moment microphysics parameterization and the RUC land-surface model; however, the PBL and radiation physics options are varied amongst the ensemble members to increase ensemble spread, given the fact that the EnKF may underrepresent model physics errors. Six-hour (three-hour) 18-member ensemble forecasts will be initialized from the real-time WoFS analyses hourly (half-hourly) from 1700 UTC Day 1 through 0300 UTC Day 2. These forecasts will be viewable using the web-based WoFS Forecast Viewer (https://wof.nssl.noaa.gov/realtime/). Table 15 shows the differences in model specifications between the HRRRE and real-time WoFS, and Figure 1 shows an example of a SPC Day 1 convective outlook and corresponding Real-time WoFS grid with WSR-88D radars used for data assimilation overlaid.

Table 15 HRRRE and Real-time WoFS configuration comparison.

	HRRRE	**Real-time WoFS**
Model Version	WRF-ARW v3.9+	WRF-ARW v3.9+
Grid Dimensions	1800 x 1060 x 50	300 x 300 x 50
Grid Resolution	3 km	3 km
EnKF cycling	36 mem. w/ GSI-EnKF **every 1 hr**	36 mem. w/ GSI-EnKF **every 15 min**
Observations	- Prepbufr conventional observations - GOES-16 ABI radiances - MRMS radar reflectivity	-Prepbufr conventional observations -Oklahoma Mesonet (when available) -MRMS reflectivity > 20 dBZ; radar 'zeroes' -MRMS radial velocity (3-km grid w/mask) -GOES-16 cloud-water path -GOES-16 clear sky radiances -GOES-16 atmospheric motion vectors -GOES-16 GLM data
Radiation LW/SW	RRTMG/RRTMG	Dudhia/RRTM, RRTMG/RRTMG
Microphysics	**Thompson (aerosol aware)**	**NSSL 2-moment**
PBL	**MYNN**	**YSU, MYJ, or MYNN**
LSM	RUC (Smirnova)	RUC (Smirnova)

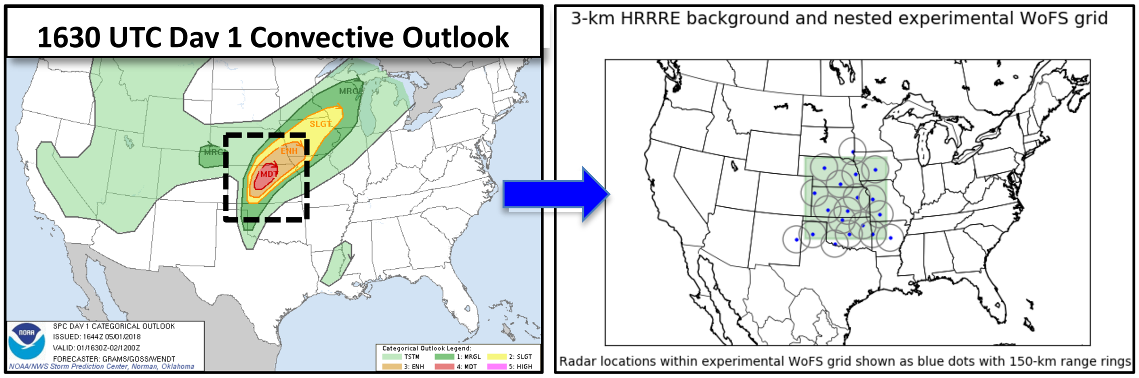

Figure 1 SPC 1630 UTC issued Day 1 convective outlook (left) and corresponding WoFS grid (right).

ii) High-resolution WoFS

To evaluate the impact of increased grid resolution on WoFS Ensemble 0-3 h forecasts, a new feature of SFE 2020 is a 9-member, 1.5 km grid-spacing WoFS Ensemble generated by downscaling from 9 members of the 3-km real-time system. The 1.5-km members are configured just like the 3-km members and will be run during the nighttime after real-time runs have finished. The 1.5 km forecasts will be initialized hourly from 1800 to 0300 UTC and will be viewable using the WoFS Forecast Viewer (https://wof.nssl.noaa.gov/realtime).

iii) Dual-resolution Hybrid data assimilation experiments

An experimental Warn-on-Forecast weather-adaptive, dual-resolution hybrid data assimilation and forecast system (WoFS-hybrid) has been developed and tested (Wang et al. 2019). The system uses WRF-ARW as its convective scale NWP model. For the data assimilation component, the system uses the combination of flow-dependent ensemble background error covariances derived from the WoFS GSI-EnKF System (Yussouf and Knopfmeier 2019; Jones et al. 2020), and static background error covariances from the NSSL variational data assimilation system (WoFS-var; Gao et al. 2013). The dual-resolution WoFS-hybrid may leverage the advantages of both WoFS-e and WoFS-var and mitigates their respective shortcomings.

In the WoFS-var component, WSR-88 radar data, satellite retrieved cloud water path, and surface observations will be used through rapid DA and forecast cycles (every 15 minutes) though some of these data will be used in different formats from those used in real-time WoFS. In addition, GOES-16 GLM data will be used in the var component (see Table 16 for details). An 18-h forecast launched from 1600 UTC with the HRRRv4 provided by GSL is used to provide boundary conditions. Similarly, a 3-h forecast launched from the 1200 UTC HRRRv4 is used to provide initial conditions for the hybrid analysis. The hybrid system will run from 1500 UTC Day 1 to 0300 UTC Day 2. A 6-h (3-h) forecast will be launched from the analysis each hour (half hour) from 1700 UTC during this period. The daily domain will be the same as the real-time WoFS. Physics specifications and how they compare with the WoFS system are provided in Table 16. These forecasts will be viewable using the web-based WoFS Forecast Viewer.

Finally, a separate WoF-var analysis and forecast system will be run similarly to the WoFS-hybrid for comparison and evaluation purposes.

Table 16 Dual resolution WoFS-hybrid configuration and comparison.

System	*WoFS*	*WoFS-hybrid*
Model Version	*WRF-ARW v3.8*	
Grid points	*300 x 300 x 50*	*600 x 600 x 50*
Grid Spacing	*3 km*	*1. 5 km*
Data Assimilation Cycle	*Every 15 min*	
Observations	*Surface data (in NCEP Burf format), Oklahoma Mesonet (when available); Doppler velocity; MRMS Radar reflectivity>20 dBZ; radar 'zeros', radial velocity and reflectivity data; cloud-water path (GOES-16)*	*NCEP Surface data (in NCEP Burf format), Oklahoma Mesonet (when available); radar raw radial velocity & reflectivity data >15 dBZ; radar 'zeros'; cloud-water path (GOES-16); GLM (lightning from GOES-16)*
PBL physics	*MYJ, YSU, or MYNN*	*MYJ*
Initialization & Boundary Conditions	*HRRRE*	*HRRR*
Factor of error covariances	*0. 5*	*0. 5*
LW/SW	*RRTMG/RRTMG*	
Microphysics	*NSSL 2-moment*	
LSM	*RUC (Smirnova)*	

e) Iowa State University (ISU) Machine Learning-based Severe Wind Probabilities

A machine-learning-based tool will be used to derive probabilities that thunderstorm wind damage reports were truly due to severe intensity winds (50 knots or more). It is well-known that there are deficiencies in the way that estimated wind values are currently assigned to thunderstorm wind damage reports. Roughly 90% of all reports do not have a measured value, and instead are given an estimate, with an artificial spike in the frequency of 50 knot and 52 knot (60 mph) values. The 50 knot estimates often appear for reports involving tree damage, implying that many of these reports are not actually due to severe intensity winds.

Several machine learning algorithms were trained on thunderstorm wind damage reports that had a measured wind value assigned to them during the 2007-2017 period. These models include a gradient boosted machine, a generalized linear model, an artificial neural network, a random forest, and an averaged ensemble. The training of these models utilized information from the Storm Report database, including textual information of the damage, along with SPC mesoanalysis output for 31 weather parameters over a 200 x 200 km box centered on the storm reports at the nearest hour prior to the report occurrence. Probabilities derived from each of these machine learning models will be available. An example is shown in Figure 2.

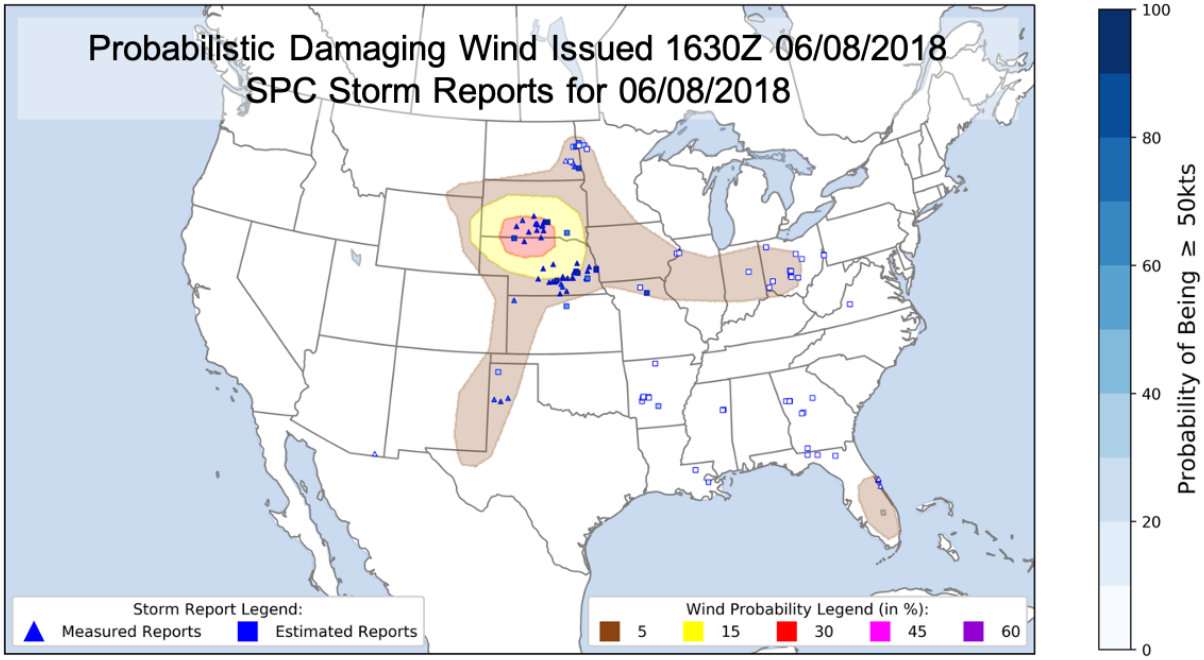

Figure 2 SPC Day 1 probabilities of damaging wind gusts (≥ 50 knots) within 40-km of a point (shaded) and corresponding observed wind reports (boxes and triangles). The color of the boxes indicates the probability that the report was associated with an actual wind gust ≥ 50 knots.

f) NCAR machine learning-derived convective hazard probabilities

For the 2020 SFE, two machine learning (ML) techniques (a random forest [RF] and neural network [NN]) are being used to produce gridded probabilistic convective hazard guidance over the contiguous United States. To do so, 38 diagnostics (Table 17) generated from 415 3-km, 36-h, 00 UTC GFS-initialized WRF forecasts of severe weather events occurring between 2010-2015 were used to train RF and NNs. The WRF diagnostics were upscaled to an 80-km grid and each grid point was labeled as a "hit" if a severe weather report occurred within a spatial and temporal neighborhood. Six labels were used, corresponding to the three report types, plus two significant report types, and a label if any report occurred. The temporal neighborhood for reports was fixed at 2-h, to produce hazard guidance within 4-h windows, while two spatial neighborhoods were tested (40 km and 120 km). This labeled set of ~10 million grid points of upscaled diagnostic fields and associated labels was used to train the RF and NNs. The configuration details of the trained ML models are provided in Table 18.

These trained models will be fixed throughout Spring 2020 and will be used to produce 4-h probabilistic convective hazard predictions of the 6 severe weather hazard types at each 80-km grid point using output from a real-time WRF forecast, configured identically to the training WRF dataset. For comparison, a smoothed mid-level updraft helicity (UH)-based forecast will also be produced, by applying a seasonally and temporally varying threshold to the maximum UH magnitude occurring within 2 h and 40 km or 120 km of a grid point. This binary field is then smoothed with a Gaussian smoother with $\boldsymbol{\sigma}$ = 160 km. The NN, RF, or UH probabilistic forecasts are all defined equivalently, i.e., probability for a particular hazard type occurring within 2 h of a forecast hour and 40 km or 120 km of a grid point.

Evaluation of the forecasts will be facilitated through a web-based visualization interface available here: https://www2.mmm.ucar.edu/projects/ncar_ensemble/camviewer/. An example 4-h all severe hazard forecast from 8 April 2020 is provided in Figure 3.

Table 17 The 42 base predictors used to train the NNs & RFs. The mean of the environmental and upper-air fields, and the maximum of the explicit fields, within each 80-km grid box, was used as input into the ML models. In addition, 132 neighborhood predictors were constructed by taking larger spatial and temporal means and maximums of the 15 environmental and 7 explicit fields resulting in a final set of 174 predictors used as input into the ML models.

Base Predictor	Type	Base Predictor	Type
Forecast Hour	Static	Surface pressure	Environment
Day of Year	Static	Most-unstable CAPE X 0-6km bulk wind difference	Environment
Latitude	Static	Significant tornado parameter	Environment
Longitude	Static	700 hPa–500 hPa lapse rate	Environment
Surface-based CAPE	Environment	Hrly-max 2–5km UH	Explicit
Most-unstable CAPE	Environment	Hrly-max 0–3km UH	Explicit
Surface-based CIN	Environment	Hrly-max 0–1km UH	Explicit
Mixed-layer CIN	Environment	Hrly-max updraft speed below 400 hPa	Explicit
0-6km bulk wind difference	Environment	Hrly-max downdraft speed below 400 hPa	Explicit
Mixed-layer lifted condensation level	Environment	Hourly-max 10-m wind speed	Explicit
0-1km bulk wind difference	Environment	Hourly precipitation accumulation	Explicit
0-1km storm-relative helicity	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa zonal wind	Upper-air
0-3km storm-relative helicity	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa meridional wind	Upper-air
2-m temperature	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa temperature	Upper-air
2-m dew point temperature	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa dew point	Upper-air

Table 18 Settings used to construct and train the NNs.

Neural Network Hyperparameter	Value
Number of hidden layers	1
Number of neurons in hidden layer	1024
Dropout rate	0.1
Learning rate	0.001
Number of training epochs	10
Hidden layer activation function	Rectified Linear Unit
Output layer activation function	Sigmoid
Optimizer	Stochastic Gradient Descent
Loss function	Binary Cross-entropy
Batch size	1024
Regularization	L2
Batch normalization	On

Table 19 Settings used to construct and train the RFs.

Random Forest Hyperparameter	Value
Number of trees	100
Maximum depth of each tree	20
Minimum number of samples to split a node	20
Minimum number of samples required to be a leaf node	10
Number of features to consider for each split	13
Splitting criterion	Gini impurity

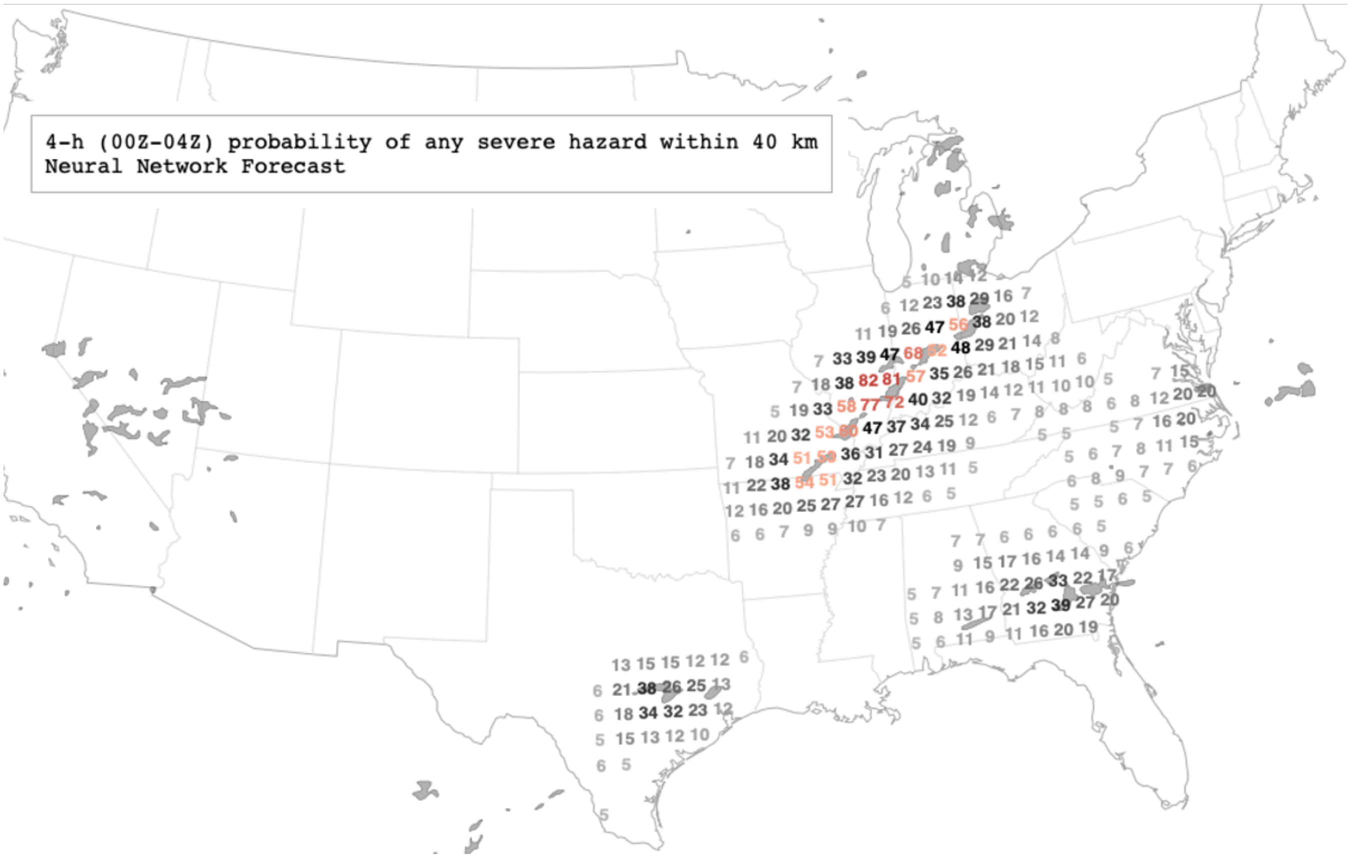

Figure 3 Neural network based probabilistic hazard forecast for the 4-h period between 00Z- 04Z 9 April 2020 based on a WRF forecast initialized at 00 UTC 8 April 2020. Numbers indicate the probability of any severe hazard occurring within 40-km of a grid point. Forecast reflectivity objects > 35 dBZ are overlaid.

g) Texas Tech University HRRRE-based Ensemble Subsetting

Ensemble sensitivity is a statistical technique applied within an ensemble that identifies features in the flow at early forecast times that are related to the predictability of chosen severe storm characteristics later in the forecast. In other words, ensemble sensitivity reveals the flow features for which associated errors will grow rapidly to adversely affect the predictive skill of chosen severe storm aspects. It can thus be expected that ensemble members that have the least error in the most sensitive regions early in a forecast window will provide better forecasts than other members, allowing the generation of adjusted and improved probabilities well before the next extended forecast cycle. *The goal of this SFE 2020 activity is to evaluate ensemble sensitivity-based subsets within the HRRRE subsets (Table 3) of the CLUE to understand whether they can improve forecasts over that from a full ensemble in a real-time environment*. The planned activity follows an evaluation at SFE 2019 of the subsetting technique which showed potential in the CLUE framework.

A daily evaluation of subset probabilities from the HRRRE against those based on the full HRRRE will be conducted. The full HRRRE is a time-lagged ensemble, with all 9 members from the 1800 and 0000 UTC initializations. The subset will be composed of 6 members chosen from the full set of 18 members. Each day, a response function location and time will be chosen through a web-based graphical user interface that identifies areas of Day 1 severe convection. The Day 1 response function will be chosen over a 6-hr period between 1800 UTC and 1200 UTC (next day) in areas where better predictions of severe convection are desired (e.g. areas of high uncertainty). Once the response time and location are chosen, the sensitivity of a single response function will be calculated: the number of grid points exceeding 50 m^2/s^2 2–5km updraft helicity. These sensitivities will be generated completely within the TTU 42-member ensemble. CLUE HRRRE members (interpolated to the TTU CONUS modeling grid) will then be chosen objectively based on their errors in the most sensitive regions using the 0600 UTC TTU ensemble analysis.

Probability fields (specifically exceedance probabilities of updraft helicity) of Day 1 convection will be generated for the 6-member HRRRE subset and will be compared against the full 18-member time-lagged HRRRE. Additionally, the HRRRE member with the lowest sensitivity-weighted errors (the "best member") will be evaluated alongside the probability fields to understand whether it regularly provides forecast value in a deterministic sense. SPC storm reports and the associated practically perfect probability field as well as MRMS data will serve as the observations against which the full and subset HRRRE probabilities and the best member are evaluated. The TTU 42-member ensemble system within which the sensitivities are generated is a DART WRF ensemble Kalman filter that assimilates numerous surface and upper-air observations on a 6-hr assimilation cycle. Assimilation is performed over a 12-km CONUS domain with downscaled 48-hr WRF forecasts run twice daily on a 4-km domain across the U.S. Midwest and South Plains. Real-time output from the Texas Tech ensemble can be viewed at http://www.atmo.ttu.edu/bancell/real_time_ENS/ttuenshome.php.

h) ML-based Severe Weather Probabilities (Loken et al. 2020)

Automated “first guess” Day 1 (1200 – 1200 UTC) hazard probabilities will be generated using machine learning (ML). For each day, eight random forests (RFs) predict, respectively, the probability of: all-hazards severe weather, all-hazards significant severe weather, any tornadoes, significant tornadoes, any severe wind, significant severe wind, any severe hail, and significant severe hail occurring within a surrounding 80 km $\times$ 80 km box. These predictions are additionally aggregated into an automated day 1 first-guess convective outlook, analogous to that produced by SPC.

The eight RFs use the same set of predictors, which are derived from temporally-aggregated (i.e., daily maximum, minimum, or mean) forecast variables (Table 20) from the 10-member HREFv2.1. Predictors include both simulated storm- and environment-related fields from each member as well as latitude and longitude. The simulated storm fields (i.e., those in the first column of Table 20) are spatially smoothed with a 2-dimensional Gaussian kernel density function before they are input into the RFs. Training occurs and output predictions are given on an approximately 80 km grid. Each RF is trained on the appropriate subset of archived observed SPC storm reports, with the training set dating back to April 2018. The methods follow those outlined in Loken et al. (2020).

Table 20 RF predictor variables. Simulated storm and environmental fields are from each member of the 10-member HREFv2. A 2-dimensional Gaussian kernel density function is applied to all storm fields before use in the RFs. Note that the RFs consider only those predictor variables located at the point of prediction.

Simulated Storm Fields	Simulated Environment Fields	Simulated Environment Fields	Miscellaneous Predictors
1 km Reflectivity (Daily Max.)	2-m Temperature (Daily Mean)	Lifted Index (Daily Min.)	Latitude
Echo Top (Daily Max.)	2-m Dewpoint Temperature (Daily Mean)	0-3 km Helicity (Daily Max.)	Longitude
Upward Vertical Velocity (Daily Max.)	Surface pressure (Daily Mean)	0-1 km Helicity (Daily Max.)	
Downward Vertical Velocity (Daily Min.)	Precipitable Water (Daily Mean)	10m-500mb Wind Shear Magnitude (Daily Mean)	
2-5 km Updraft Helicity (Daily Max.)	Hourly Max. 10-m Wind Components (Daily Mean)	10m-500mb Wind Shear Direction (Daily Mean)	
0-3 km Updraft Helicity (Daily Max.)	MUCAPE (Daily Mean)	10m-925mb Wind Shear Magnitude (Daily Mean)	
	MUCIN (Daily Mean)	Critical Angle Proxy (Daily Mean)	
	0-6 km Storm Relative Motion Components (Daily Mean)	CAPE × 10m-500mb Wind Shear (Daily Mean)	

3. SFE 2020 Core Interests and Daily Activities

Model evaluation and forecasting activities will occur virtually 10am–noon and 1:30–4pm, respectively. Tables 21 & 22 provide a schedule; further details are provided in subsequent sections.

Table 21 Model evaluations schedule.

Model Evaluations: Monday		
Time (CDT)	*Topic*	*Moderator*
10:00 a.m.	**Welcome and Introductions**	Israel
10:20 a.m.	**Overview of SFE Model Contributions and Scientific Goals**	Israel and PIs
11:00 a.m.	**Preview of the Evaluations** (Science Questions, Examples)	Group A: Israel & David J. Group B: Burkely & Adam
Model Evaluations: Tuesday-Friday		
9:45 a.m.	**Overview of Yesterday's Severe Weather** (David Imy) Break into Virtual Groups (A & B)	
	Group A (Israel & David J.)	***Group B (Burkely & Adam)***
10:00 a.m.	**Independent Evaluations** (with moderators available for questions)	**Independent Evaluations** (with moderators available for questions)
11:00 a.m.	**Discussion of Evaluations:** A1. ISU ML Severe Wind Probs A2. NCAR ML Hazard Guidance A3. CLUE: 00Z CAM TL-Ensemble A4. CLUE: TTU Ensemble Subsetting A5. CLUE: Ens. Hail Guidance *(Fri)* A6. CLUE: FV3-SAR Physics/DA/VL A7. CLUE: FV3-SAR IC/Hord/LSM A8. Mesoscale Analyses A9. CLUE: Lightning DA	**Discussion of Evaluations:** B1(a-f). HREF Calibrated Guidance B2. CLUE: 00Z CAM Multi-Model Ens. B3. CLUE: 12Z CAM TL-Ensemble B4. CLUE: Deterministic Flagships B5. CLUE: Core and ICs B6(a-f). WoFS Configurations

Table 22 Short-term forecasting schedule.

Short-Term Forecasting: Monday-Friday		
1:30 p.m.	**Overview of Today's Severe Weather Threat** (David Imy) Break into Virtual Groups (R2O & Innovation)	
	R2O (Israel & Mike)	***Innovation (David Imy & Adam)***
1:40 p.m.	**Overview of SFE Drawing Tool (M); Evaluation of Yesterday's Forecasts (T-F)**	**Overview of WoFS Drawing Tool (M); Evaluation of Yesterday's Forecasts (T-F)**
2:00 p.m.	**Day 1 Outlook Generation*** Full period (20-12Z) coverage and conditional intensity forecasts of tornado, hail, and wind using available 12Z CAM ensemble guidance (not WoFS) and observations.	**Short-Term Outlook Generation*^** 1-h (21-22Z) and 4-h (21-01Z) probabilistic forecasts of tornado, hail, and wind. Some forecasters with access to WoFS^ and some without*.
3:00 p.m.	**Day 1 Outlook Update*** Update full period (21-12Z) coverage and conditional intensity forecasts of tornado, hail, and wind using WoFS and observations.	**Short-Term Outlook Update*^** 1-h (21-22Z) and 4-h (21-01Z) probabilistic forecasts of tornado, hail, and wind. Same forecasters with access to WoFS^ and same without*.

* Using SFE Drawing Tool
^ Using WoFS Drawing Tool

a. Formal Evaluation Activities

SFE 2020 will feature one period of formal evaluation from 10am-noon. The evaluations will be done virtually and involve comparisons of different ensemble diagnostics, CLUE ensemble subsets, HREF, and the WoFS Ensembles. Participants will be split into Groups A & B, which will each conduct a separate set of evaluations. On Mondays, SFE coordinators will go through the evaluations with participants to provide background information and instructions. Then, Tuesday – Friday, participants will conduct the evaluations individually during the 10-11am time period. Then, from 11am-noon, each group will reconvene in a virtual meeting (Google Meet) to discuss various aspects of the evaluations (e.g., questions anybody has, interesting differences, particularly notable differences in performance, etc.). The two different sets of evaluations are summarized below.

Group A

A1. ISU ML Severe Wind Probs

An evaluation will be conducted of a ML-based tool to estimate the likelihood that a damaging wind report was caused by wind ≥ 50 knots. The evaluations will focus on perceived usefulness of the output via comparison with SPC forecasts of severe wind probability, best methods to display the information, and subjective evaluation of different ML techniques. The evaluation will be conducted on an external web page hosted by Iowa State University.

Primary Science Question: Can machine-learning approaches provide useful information regarding the likelihood of wind damage reports being associated with gusts ≥ 50 knots?

A2. NCAR ML Hazard Guidance

Two ML techniques (random forest and neural network) will be used to produce gridded probabilistic hazard guidance using a deterministic 3-km grid-spacing WRF model configuration run by NCAR. Evaluation will focus on the perceived utility of the ML-based based products and subjective evaluation of the two unique ML techniques. The evaluation will be conducted on an external web page hosted by NCAR https://www2.mmm.ucar.edu/projects/ncar_ensemble/camviewer/.

Primary Science Question: Can machine-learning approaches provide guidance that can discriminate among severe weather hazards?

A3. CLUE: 00Z CAM TL-Ensemble

This evaluation will compare two single-model ensembles (HRRRE and UM) initialized at 0000 UTC with their respective time-lagged ensembles (I.e., half of the members each from 0000 UTC and 1800 UTC). These single-model ensembles will be compared to the HREFv2.1 and HREFv3, which serve as the baseline for CAM ensemble performance. The goal is to assess whether time-lagging results in improved probabilistic forecasts from single-model ensembles.

Primary Science Question: Is time-lagging a viable option for formal CAM ensemble design?

A4. CLUE: TTU Ensemble Subsetting

Severe weather probabilities for a 6-member HRRRE subset will be compared to those from the "full" HRRRE composed of the 18-members initialized from 1800 and 0000 UTC. The 6-member subset is composed of the members with the smallest errors in sensitive regions as determined by ensemble sensitivity analysis. Additionally, the best HRRRE member (i.e., lowest sensitivity-weighted errors) will also be evaluated.

Primary Science Question: Can a sensitivity-based ensemble subsetting approach lead to improved guidance over the full ensemble for severe-weather forecasting?

A5. CLUE: Ens. Hail Guidance (Fridays only)

Maximum hail size fields will be formally evaluated within the HRRRE. These hail size forecasts will include those derived by (1) the HAILCAST algorithm (Adams-Selin and Ziegler 2016), which predicts maximum hail size using a hail growth model coupled to WRF, (2) the Thompson method, which estimates hail size directly from the microphysics size distribution by finding the largest graupel or hail hydrometeor diameter that exceeds a specified number concentration, (3) neighborhood-based, probabilistic forecasts of UH exceeding a fixed threshold loosely calibrated to maximize the fractions skill score (FSS) for 1-inch hail, and (4) a machine-learning-based method that provides probabilistic hail size forecasts (Gagne et al. 2017). Comparisons will be made to hail LSRs and MRMS MESH.

Primary Science Question: What is the best algorithm/approach to estimate hail size from CAM ensembles?

A6. CLUE: FV3-SAR Physics/DA/VL

This comparison (4-panel display) will examine the FV3-SAR configurations provided by EMC (Table 11) that vary the physics suite and add a data assimilation component. In addition, a similarly configured FV3-SAR run from NSSL (Table 10) with additional vertical levels will be examined. Particular attention will be given to simulated storm structure, convective evolution, and location/coverage of storms, as well the impact of these configurations on 2-m T/Td and thermodynamic sounding profiles.. Storm surrogate fields, like hourly maximum updraft helicity, will also be examined to gauge their utility for forecasting severe storms.

Primary Science Question: What is the optimal configuration of the FV3-SAR for convective weather forecasting?

A7. CLUE: FV3-SAR IC/Hord/LSM

This set of comparisons (6-panel display) will examine the set of FV3-SAR runs provided by GSL (Table 4), along with an FV3-SAR run provided by EMC and initialized from GFS (Table 11). The goal of this evaluation is to compare the relative impact of initial conditions, land surface model, and horizontal advection settings on the forecasts. Particular attention will be given to simulated storm structure, convective evolution, and location/coverage of storms, as well the impact of these configurations on 2-m T/Td and thermodynamic sounding profiles.

Primary Science Question: What is the optimal configuration of the FV3-SAR for convective weather forecasting?

A8. Mesoscale Analyses

This evaluation will examine 15-minute output of two different versions of the 3D real-time mesoscale analysis (3D-RTMA) system that use HRRRv4 forecasts as the background. One version provided by GSL uses the GDAS for background error covariance information in the hybrid DA system, while the other version provided by EMC uses the HRRRDAS for background error covariance information. The goal is to assess the utility of these analysis systems for situational awareness and short-term forecasting for convective-weather scenarios.

Primary Science Question: Does using information from a CAM ensemble in the hybrid ensemble-variational analysis improve its utility for short-term convective forecasting applications?

A9. CLUE: Lightning DA

This comparison will focus on forecast hours 0-12 over areas with sparse radar coverage, so the focus areas may be different from that of the other experiments focused on severe weather. WRF-ARW simulations with HRRRv4 physics that use a convective scale data assimilation system, NEWS3DVAR (Fierro et al. 2016, 2018), with and without assimilation of total lightning data from the GOES 16 Geostationary Lightning Mapper (GLM) will be examined (Table 8) using a 3-panel display. Particular attention will be given to location/coverage of storms, simulated storm structure, and convective evolution., and location/coverage of storms.

Primary Science Question: Does the assimilation of GOES-16 GLM data improve short-term forecasts (0-12 hours) of thunderstorms in radar-sparse regions?

Group B

B1. HREF Calibrated Guidance

a. 24-h tornado guidance

Climatological frequencies of tornadoes associated with specified ranges of STP are used to derive tornado probabilities using simulated updraft helicity and STP for 40-km radius circular areas (Gallo et al. 2018; STP Cal. Circle). Two other similar approaches are used that only consider STP within storm inflow quadrants – one that attempts to filter out UH associated with MCSs, and one without the MCS filter. Additionally, calibrated probabilities using a combination of SREF and HREF data are evaluated (Jirak et al. 2014), along with probabilities derived from the Loken et al. (2020) random forest machine learning algorithm.

b. 4-h tornado guidance

For 4-h tornado probabilities, only the STP Cal. Circle and HREF/SREF calibrated probabilities are evaluated along with SPC Timing Guidance (derived from SPC Outlooks and HREF/SREF calibrated guidance using a temporal disaggregation approach; Jirak et al. 2012; Jirak et al. 2020).

c. 24-h hail guidance

Machine learning-based forecasts using the methods outlined in Burke et al. (2020) are evaluated along with HREF/SREF calibrated probabilities, and Loken et al. (2020) machine-learning-based probabilities.

d. 4-h hail guidance

Burke et al. (2020) ML forecasts, temporally disaggregated SPC Timing Guidance, and HREF/SREF calibrated probabilities are examined.

e. 24-h wind guidance

HREF/SREF and Loken et al. (2020) calibrated probabilities are examined.

f. 4-h wind guidance

HREF/SREF calibrated probabilities and SPC Timing Guidance are examined.

Primary Science Question: What are the best approaches and techniques to develop calibrated hazard probabilities from the HREF?

B2. CLUE: 00Z CAM Multi-Model Ens.

This evaluation will compare single-model, 18-member time lagged HRRRE and UM ensembles, while another comparison will evaluate 0000 UTC multi-model (HRRRE and UM), 18-member single model ensembles and a 36-member, time-lagged, multi-model ensemble. HREF (v2.1 and v3) will be used as a baseline for performance. The goal is to explore the sensitivity of time-lagging and multi-model strategies in CAM ensemble configurations.

Primary Science Question: What is the relative impact of time-lagging and multi-model approaches in CAM ensembles for next-day severe-weather forecasting?

B3. CLUE: 12Z CAM TL-Ensemble

Three nine-member, single-model ensembles based at 1200 UTC that use different time-lagging strategies will be compared. The first ensemble, HRRRE, does not use time-lagging; the second ensemble, HRRRE-TL9, uses 3 members from the 0000, 0600, and 1200 UTC HRRRE initializations; and the third ensemble, HRRR/NSSL-WRF-TL, uses a mix of five HRRR and four NSSL-WRF configurations each initialized with HRRRv4 initial conditions at different times between 0000 and 1200 UTC. The goal is to determine whether time-lagged ensembles (of multiple designs) can provide improved probabilistic guidance over a single-model ensemble initialized at a single time.

Primary Science Question: What is the optimal design for a single-model, time-lagged ensemble?

B4. Deterministic Flagships

This activity will focus on assigning ratings to gauge the skill and utility of the primary deterministic CAMs provided by each SFE collaborator – UK Met Office (*um-ICs02*), GFDL (*gfdl-fv3*), NSSL (*sarfv3-ICs01*), and EMC (*emc-fv3sar*). These runs will be compared to the HRRRv4, which was developed by GSL and will soon be operational. Particular attention will be given to simulated storm structure, convective evolution, and location/coverage of storms. Storm surrogate fields, like hourly maximum updraft helicity, will also be examined to gauge their utility for forecasting severe storms.

Primary Science Question: How do the deterministic CAM runs using the UM and FV3 dynamic core compare to the operational standard for convective forecasting (i.e., WRF-ARW)?

B5. CLUE: Core and ICs

In this comparison, a matrix of 6 simulations using multiple model cores: WRF-ARW (Table 5), SAR-FV3-SAR (Table 10), and UM (Table 7), each initialized from GFS and UK Met Office global models, will be examined. Each row of panels will display a single model core with corresponding observations to the right, and each column will display the same set of initial conditions. The goal is to assess the sensitivity of convection-allowing model runs to initial conditions and model core. Particular attention will be given to simulated storm structure, convective evolution, and location/coverage of storms. Storm surrogate fields, like hourly maximum updraft helicity, will also be examined to gauge their utility for forecasting severe storms.

Primary Science Question: What is the relative impact of initial conditions and model core on next-day convective forecasts?

B6. WoFS Configurations

a.-c. Real-time 3-km vs. Experimental 1.5-km

This comparison will examine neighborhood maximum ensemble probabilities of storm diagnostics like updraft helicity, updraft speed, 10-m wind speed, and hail size (HAILCAST) compared to

LSRs for tornado, hail and wind, MRMS reflectivity and MESH, and NWS warnings. A set of 2-panels will display each WoFS configuration for (a) 2000, (b) 2200, and (c) 0000 UTC WoFS initializations.

d. Evolution of forecasts with decreasing lead time

In this 6-panel comparison, 1700, 1800, 1900, 2000, 2100, and 2200 UTC initializations of the real-time WoFS configuration valid for the same time window, will be examined to gauge the evolution of the WoFS forecasts with decreasing lead time.

e.-f. *Hybrid and Var DA comparison*

These comparisons will examine the deterministic 1.5 km grid-spacing Hybrid and Var data assimilation runs (Table 16) initialized from (a) 2000, and (b) 2200 UTC.

Primary Science Question: Does reduced grid spacing improve WoFS ability to forecast convective storm mode, evolution, and intensity?

b. Forecast products and activities

Because of the COVID-19 pandemic, forecasting activities will be limited in scope and occur virtually from 1:30-4pm daily with a focus on adding temporal specificity to convective outlooks within the Day 1 time period using WoFS Ensemble datasets. Participation will be limited to a small internal group, as well as weekly groups of NWS forecasters. The experimental forecasts will cover a limited-area domain typically covering the primary severe threat area with a domain based on existing SPC outlooks and/or where interesting convective forecast challenges are expected. As in previous years, two sets of unique outlooks will be generated by the R2O and Innovation desks. Both desks will issue outlooks for the probability of individual hazards (tornado, wind gusts $\geq$ 50 knots, hail $\geq$ 1.0 in.) within 25 miles (40 km) of a point. For the Innovation desk, forecasting individual hazards is a shift from previous years when all the outlooks were focused on total severe (i.e., all hazards combined).

At the R2O desk, participants will update the operational SPC 1630Z Day 1 Outlook hazard probabilities for the period 2100 – 1200 UTC. Additionally, conditional intensity forecasts will be generated, for which SPC's operational probabilities of significant severe hazards (EF-2 or greater tornadoes, winds $\geq$ 65 kts, or hail $\geq$ 2 in.) can be used as a starting point. This will be the second year that the R2O desk has issued conditional intensity forecasts. These forecasts delineate areas that are expected to follow a "normal", "hatched", or "double-hatched" distribution. In plain language, "normal" refers to a typical severe weather day, where significant severe weather is unlikely, "hatched" areas indicate where significant severe weather is possible, and "double-hatched" areas indicate where high-impact significant severe weather is expected. These forecasts could also be thought of as indicating the proportion of observed reports that are expected to be severe, where going from "normal" to, "hatched" to , and "double-hatched" would indicate an increasing proportion of significant-severe reports. One set of forecasts will be generated 2-3pm for which WoFS data will not be used, and a final set of forecasts will be generated 3-4pm with available WoFS datasets. All forecasts at the R2O desk will be generated using the "SFE viewer" (https://hwt.nssl.noaa.gov/sfe_viewer/2020/forecast_tool).

At the Innovation desk, participants will generate severe hazard probabilities valid over a short time window, 4-5pm (2100-2200 UTC), and a long time window, 4-8pm (2100-0100 UTC). An initial forecast will be generated during the 2-3pm period and an updated final forecast during the 3-4pm period. For both sets of initial and final forecasts, one group of forecasters will be able to use all available datasets except for WoFS, while another will use all available datasets including WoFS. The group that doesn't use WoFS will use the SFE viewer to generate forecasts, while the groups that uses WoFS will use the WoFS viewer (https://wof.nssl.noaa.gov/research/).

These WoF activities are the fourth year the WoF Ensemble has been tested in the EFP, and explores the potential utility of WoF products for issuing guidance between the watch and warning time scales (i.e. 0.5 to 6-h lead times). These activities represent efforts to explore ways of seamlessly merging probabilistic severe weather outlooks with probabilistic severe weather warnings as part of NOAA's Warn-on-Forecast (WoF; Stensrud et al. 2009) and Forecasting a Continuum of Environmental Threats (FACETs; Rothfusz et al. 2018) initiatives. These efforts also support the transition to higher temporal resolution forecasts at the SPC.

Appendix A: List of scheduled SFE 2020 participants.

Week 1 April 27-May 1	Week 2 May 4-8	Week 3 May 11-15	Week 4 May 18-22	Week 5 May 26-29
Lizzie Tirone (ISU)	Lizzie Tirone (ISU)	Lizzie Tirone (ISU)	Lizzie Tirone (ISU)	Lizzie Tirone (ISU)
Jeremiah Pyle (AWC)	Brice Coffer (NCState)	Bill Gallus (ISU)	Lance Bosart (Suny-Albany)	Kallan Parker (PSU; Hollings)
Victor Gensini (NIU)	Lucia Scaff (U. Sask)	Kyle Hugeback (ISU)	Bruno Ribeiro (Suny-Albany)	Clark Evans (UWM)
Ryan Sobash (NCAR)	Corey Potvin (NSSL)	Michou Baart de la Faille (KNMI)	Scott Feldman (Suny-Albany)	Dillon Blount (UWM)
Yongming Wong (OU/MAP)	Becky Adams-Selin (AER)	Tina Kalb (DTC)	Steve Weiss (SPC Ret.)	Craig Schwartz (NCAR)
Amanda Burke (OU)	Alicia Bentley (EMC)	John Allen (CMU)	Harald Richter (BoM)	Ben Blake (EMC)
Jacob Carley (EMC)	Aaron Johnson (MAP)	Glen Romine (NCAR)	Tom Galarneau (CIMMS/NSSL)	Xiaoyan Zhang (EMC)
Brett Borchardt (WFO LOT)	Andrew McKaughan (WFO PIH)	Paige Crafter (USAF)	Tony Oakley (USAF)	Austin Coleman (TTU)
Matt Anderson (WFO MRX)	Alex Lukinbeal (WFO MSO)	Logan Dawson (EMC)	Gang Zhou (EMC)	Jidong Gao (NSSL)
Alex Krull (WFO DMX)	Hayden Frank (WFO BOX)	Austin Dixon (OU)	Matt Pyle (EMC)	Jamie Wolff (DTC)
David Harrison (CIMMS/SPC)	Patrick Skinner (CIMMS/NSSL)	Austin Coleman (TTU)	Austin Coleman (TTU)	Corey Mead (WFO OAX)
Derek Stratman (CIMMS/NSSL)	Yibing Su (Princeton)	Mike Seaman (WFO SLC)	Jason Godwin (WFO FWD)	Nick Vertz (WFO BYZ)
Joe Pollina (WFO OKX)	Jeff Beck (GSL)	Eric Bunker (WFO TAE)	Tom Hultquist (WFO MPX)	Curtis Alexander (GSL)
Jeff Duda (GSL)	Terra Ladwig (GSL)	Robert Megnia (WFO LCH)	Dan McKemy (WFO LMK)	John Brown (GSL)
Dave Turner (GSL)	Nigel Roberts (UK Met)	Steve Zubrick (WFO LWX)	Mike Evans (WFO ALY)	
Aurore Porson (UK Met)		Geoff Manikin (EMC)	David Dowell (GSL)	
		John Brown (GSL)	Eric James (GSL)	
		Ed Szoke (GSL)	Mike Bush (UK Met)	
		Aurore Porson (UK Met)	Dave Ahijevych (NCAR)	
		Nigel Roberts (UK Met)		
		Bethany Earnest (CIMMS/SPC)		

SFE Facilitators: Adam Clark (NSSL), Israel Jirak (SPC), Dave Imy (retired SPC), Mike Coniglio (NSSL/SPC), Burkely Gallo (CIMMS/SPC), Kenzie Krocak (CIMMS/NSSL/OU), Brett Roberts (CIMMS/SPC/NSSL), Kent Knopfmeier (CIMMS/NSSL), and Andy Dean (SPC).

Appendix B: Organizational structure of the NOAA/Hazardous Weather Testbed

NOAA's Hazardous Weather Testbed (HWT) is a facility jointly managed by the National Severe Storms Laboratory (NSSL), the Storm Prediction Center (SPC), and the NWS Oklahoma City/Norman Weather Forecast Office (OUN) within the National Weather Center building on the University of Oklahoma South Research Campus. The HWT is designed to accelerate the transition of promising new meteorological insights and technologies into advances in forecasting and warning for hazardous mesoscale weather events throughout the United States. The HWT facilities are situated between the operations rooms of the SPC and OUN. The proximity to operational facilities, and access to data and workstations replicating those used operationally within the SPC, creates a unique environment supporting collaboration between researchers and operational forecasters on topics of mutual interest.

The HWT organizational structure is composed of three overlapping programs (Fig. B1). The Experimental Forecast Program (EFP) is focused on predicting hazardous mesoscale weather events on time scales ranging from hours to a week in advance, and on spatial domains ranging from several counties to the CONUS. The EFP embodies the collaborative experiments and activities previously undertaken by the annual SPC/NSSL Spring Experiments. For more information see http://www.nssl.noaa.gov/projects/hwt/efp/.

The Experimental Warning Program (EWP) is concerned with detecting and predicting mesoscale and smaller weather hazards on time scales of minutes to a few hours, and on spatial domains from several counties to fractions of counties. The EWP embodies the collaborative warning-scale experiments and technology activities previously undertaken by the OUN and NSSL. For more information about the EWP see http://www.nssl.noaa.gov/projects/hwt/ewp/. A key NWS strategic goal is to extend warning lead times through the "Warn-on-Forecast" concept (Stensrud et al. 2009),

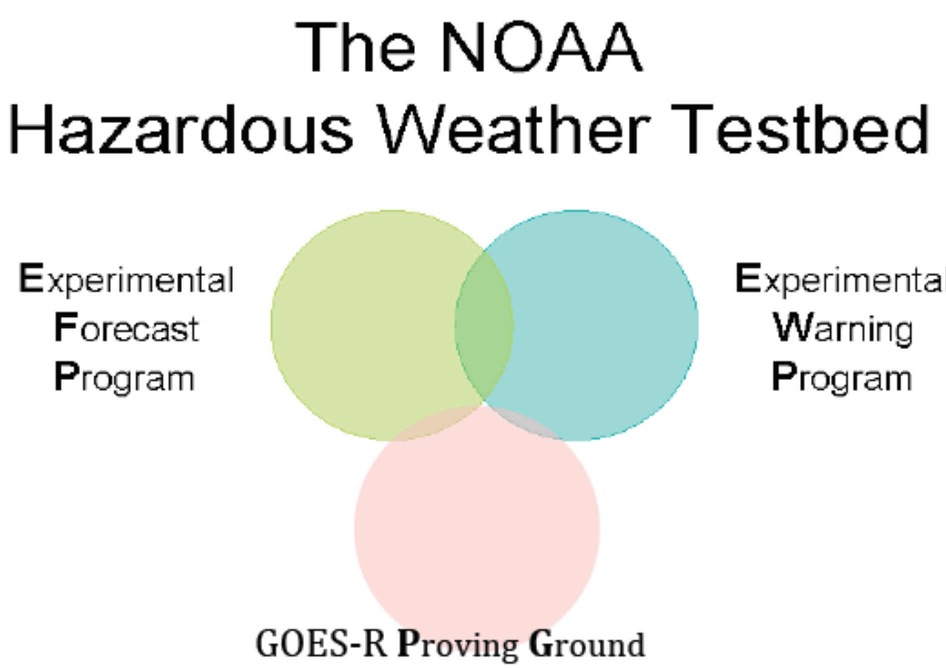

Figure B1: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP), the Experimental Warning Program (EWP), and the GOES-R Proving Ground (GOES-R).

which involves using frequently updated short-range forecasts (≤ 1h lead time) from convection-resolving ensembles. This provides a natural overlap between the EFP and EWP activities.

The GOES-R Proving Ground (established in 2009) exists to provide demonstration of new and innovative products as well as the capabilities available on the next generation GOES-16 satellite. The PG interacts closely with both product developers and NWS forecasters. More information about GOES-R Proving Ground is found at http://cimss.ssec.wisc.edu/goes_r/proving-ground.html.

Rapid science and technology infusion for the advancement of operational forecasting requires direct, focused interactions between research scientists, numerical model developers, information technology specialists, and operational forecasters. The HWT provides a unique setting to facilitate such interactions and allows participants to better understand the scientific, technical, and operational challenges associated with the prediction and detection of hazardous weather events. The HWT allows participating organizations to:

- Refine and optimize emerging operational forecast and warning tools for rapid integration into operations
- Educate forecasters on the scientifically correct use of newly emerging tools and to familiarize them with the latest research related to forecasting and warning operations
- Educate research scientists on the operational needs and constraints that must be met by any new tools (e.g., robustness, timeliness, accuracy, and universality)
- Motivate other collaborative and individual research projects that are directly relevant to forecast and warning improvement

For more information about the HWT, see http://www.nssl.noaa.gov/hwt/. Detailed historical background about the EFP Spring Experiments, including scientific and operational motivation for the intensive examination of high resolution NWP model applications for convective weather forecasting, and the unique collaborative interactions that occur within the HWT between the research and operational communities, are found in Weiss et al. (2010 – see http://www.spc.noaa.gov/publications/weiss/hwt-2010.pdf), Clark et al. (2012; 2018), and Gallo et al. (2017).

Appendix C: CLUE system output fields

Table C1 The minimum set of 123 output diagnostics for the WRF-ARW CLUE members, which are output at hourly intervals.

Number	Level/Layer	Parameter	Description
001	entire atmosphere	REFC	Composite reflectivity [dB]
002	cloud top	RETOP	Echo Top [m]
003	entire atmosphere	VIL	Radar-Simulated Vertically Integrated Liquid [kg/m^2]
004	surface	VIS	Visibility [m]
005	1000 m above ground	REFD	Reflectivity [dB]
006	4000 m above ground	REFD	Reflectivity [dB]
007	surface	GUST	Wind Speed (Gust) [m/s]
008	500 mb	HGT	Geopotential Height [gpm]
009	500 mb	TMP	Temperature [K]
010	500 mb	DPT	Dew Point Temperature [K]
011	500 mb	UGRD	U-Component of Wind [m/s]
012	500 mb	VGRD	V-Component of Wind [m/s]
013	700 mb	HGT	Geopotential Height [gpm]
014	700 mb	TMP	Temperature [K]
015	700 mb	DPT	Dew Point Temperature [K]
016	700 mb	UGRD	U-Component of Wind [m/s]
017	700 mb	VGRD	V-Component of Wind [m/s]
018	850 mb	HGT	Geopotential Height [gpm]
019	850 mb	TMP	Temperature [K]
020	850 mb	DPT	Dew Point Temperature [K]
021	850 mb	UGRD	U-Component of Wind [m/s]
022	850 mb	VGRD	V-Component of Wind [m/s]
023	925 mb	TMP	Temperature [K]
024	925 mb	DPT	Dew Point Temperature [K]
025	925 mb	UGRD	U-Component of Wind [m/s]
026	925 mb	VGRD	V-Component of Wind [m/s]
027	1000 mb	TMP	Temperature [K]
028	1000 mb	DPT	Dew Point Temperature [K]
029	1000 mb	UGRD	U-Component of Wind [m/s]
030	1000 mb	VGRD	V-Component of Wind [m/s]
031	100-1000 mb above ground	MAXUVV	Hourly Max upward Vertical Velocity - lowest 100hPa [m/s]

032	100-1000 mb above ground	MAXDVV	Hrly Max downward Vertical Velocity - lowest 100hPa [m/s]
033	0.5-0.8 sigma layer	DZDT	Vertical Velocity (Geometric) [m/s]
034	mean sea level	PRMSL	Pressure Reduced to MSL [Pa]
035	1000 mb	HGT	Geopotential Height [gpm]
036	1000 m above ground	MAXREF	Hourly Max of Simulated Reflectivity at 1 km AGL [dB]
037	5000-2000 m above ground	MXUPHL	Hrly Max Updraft Helicity - 2km to 5 km AGL [m^2/s^2]
038	entire column	TCOLG	Total Column Integrated Graupel [kg/m^2]
039	surface	LTNG	Lightning [non-dim]
040	80 m above ground	UGRD	U-Component of Wind [m/s]
041	80 m above ground	VGRD	V-Component of Wind [m/s]
042	surface	PRES	Pressure [Pa]
043	surface	HGT	Geopotential Height [gpm]
044	surface	TMP	Temperature [K]
045	0 m underground	MSTAV	Moisture Availability [%]
046	surface	WEASD	Water Equivalent of Accumulated Snow Depth [kg/m^2]
047	surface	SNOWC	Snow Cover [%]
048	surface	SNOD	Snow Depth [m]
049	2 m above ground	TMP	Temperature [K]
050	2 m above ground	SPFH	Specific Humidity [kg/kg]
051	2 m above ground	DPT	Dew Point Temperature [K]
052	10 m above ground	UGRD	U-Component of Wind [m/s]
053	10 m above ground	VGRD	V-Component of Wind [m/s]
054	10 m above ground	WIND	Wind Speed [m/s]
055	surface	CPOFP	Percent frozen precipitation [%]
056	surface	PRATE	Precipitation Rate [$kg/m^2/s$]
057	surface	APCP	Total Precipitation [kg/m^2]
058	surface	WEASD	Water Equivalent of Accumulated Snow Depth [kg/m^2]
059	surface	APCP	Precipitation [kg/m^2] – hourly total
060	surface	WEASD	Water Equivalent of Accumulated Snow Depth [kg/m^2]
061	surface	CSNOW	Categorical Snow [-]
062	surface	CICEP	Categorical Ice Pellets [-]
063	surface	CFRZR	Categorical Freezing Rain [-]

064	surface	CRAIN	Categorical Rain [-]
065	surface	VGTYP	Vegetation Type [Integer(0- 13)]
066	500-1000 mb	LFTX	Surface Lifted Index [K]
067	surface	CAPE	Convective Available Potential Energy [J/kg]
068	surface	CIN	Convective Inhibition [J/kg]
069	entire column	PWAT	Precipitable Water [kg/m^2]
070	low cloud layer	LCDC	Low Cloud Cover [%]
071	middle cloud layer	MCDC	Medium Cloud Cover [%]
072	high cloud layer	HCDC	High Cloud Cover [%]
073	entire atmosphere	TCDC	Total Cloud Cover [%]
074	cloud base	PRES	Pressure [Pa]
075	cloud base	HGT	Geopotential Height [gpm]
076	cloud ceiling	HGT	Geopotential Height [gpm]
077	cloud top	PRES	Pressure [Pa]
078	cloud top	HGT	Geopotential Height [gpm]
079	top of atmosphere	ULWRF	Upward Long-Wave Rad. Flux [W/m^2]
080	surface	DSWRF	Downward Short-Wave Radiation Flux [W/m^2]
081	3000-0 m above ground	HLCY	Storm Relative Helicity [m^2/s^2]
082	1000-0 m above ground	HLCY	Storm Relative Helicity [m^2/s^2]
083	0-6000 m above ground	USTM	U-Component Storm Motion [m/s]
084	0-6000 m above ground	VSTM	V-Component Storm Motion [m/s]
085	0-1000 m above ground	VUCSH	Vertical U-Component Shear [1/s]
086	0-1000 m above ground	VVCSH	Vertical V-Component Shear [1/s]
087	0-6000 m above ground	VUCSH	Vertical U-Component Shear [1/s]
088	0-6000 m above ground	VVCSH	Vertical V-Component Shear [1/s]
089	180-0 mb above ground	4LFTX	Best (4 layer) Lifted Index [K]
090	180-0 mb above ground	CAPE	Convective Available Potential Energy [J/kg]
091	180-0 mb above ground	CIN	Convective Inhibition [J/kg]
092	surface	HPBL	Planetary Boundary Layer Height [m]
093	lifted condensation level	HGT	Geopotential Height [gpm]
094	90-0 mb above ground	CAPE	Convective Available Potential Energy [J/kg]
095	90-0 mb above ground	CIN	Convective Inhibition [J/kg]
096	255-0 mb above ground	CAPE	Convective Available Potential Energy [J/kg]
097	255-0 mb above ground	CIN	Convective Inhibition [J/kg]
098	equilibrium level	HGT	Geopotential Height [gpm]
099	255-0 mb above ground	PLPL	Pressure of level from which parcel was lifted [Pa]

100	surface	LAND	Land Cover (0=sea, 1=land) [Proportion]
101	surface	ICEC	Ice Cover [Proportion]
102	250 mb	UGRD	U-component of wind [m/s]
103	250 mb	VGRD	V-component of wind [m/s]
104	250 mb	HGT	Geopotential Height [gpm]
105	250 mb	TMP	Temperature [K]
106	700 mb	VVEL	Vertical Velocity [m/s]
107	-10 C	REFD	Reflectivity [dB]
108	-10 C	REFD	Hourly maximum of -10C reflectivity
109	5000-2000 m above ground	MNUPHL	Hrly Min Updraft Helicity - 2km to 5 km AGL [m^2/s^2]
110	2000-0 m above ground	MXUPHL	Hrly Max Updraft Helicity - 0km to 2 km AGL [m^2/s^2]
111	2000-0 m above ground	MNUPHL	Hrly Min Updraft Helicity - 0km to 2 km AGL [m^2/s^2]
112	3000-0 m above ground	MXUPHL	Hrly Max Updraft Helicity - 0km to 3 km AGL [m^2/s^2]
113	3000-0 m above ground	MNUPHL	Hrly Min Updraft Helicity - 0km to 3 km AGL [m^2/s^2]
114	2000-0 m above ground	RELV	Hrly Max Rel. Vort. – 0km to 2km AGL [1/s]
115	1000-0 m above ground	RELV	Hrly Max Rel. Vort. – 0km to 1km AGL [1/s]
116	entire column	HAIL	Hrly Max of Hail/Graupel Diameter [m]
117	0.1 sigma	HAIL	Hrly Max of Hail/Graupel Diameter [m]
118	5000-2000m AGL	UPHL	Updraft Helicity (instantaneous) [m^2/s^2]
119	6000-1000m AGL	UPHL	Updraft Helicity (instantaneous) [m^2/s^2]
120	top of atmos	SBT123	Simulated Brightness T for GOES 12 Ch. 3 [K]
121	top of atmos	SBT124	Simulated Brightness T for GOES 12 Ch. 4 [K]
122	top of atmos	SBT113	Simulated Brightness T for GOES 11 Ch. 3 [K]
123	top of atmos	SBT114	Simulated Brightness T for GOES 11 Ch. 4 [K]

Appendix D: References

Adams-Selin, R.D. and C.L. Ziegler, 2016: Forecasting Hail Using a One-Dimensional Hail Growth Model within WRF. *Mon. Wea. Rev.,* **144**, 4919–4939,https://doi.org/10.1175/MWR-D-16-0027.1

Burke, A., N. Snook, D. J. Gagne, S. McCorkle, A. McGovern, 2020: Calibration of Machline Learning-Based Probabilistic Hail Predictions for Operational Forecasting, *Wea. Forecasting*, **35**, 149-168.

Chen, J-H., and Coauthors, 2019: Advancements in Hurricane Prediction with NOAA's Next Generation Forecast System. Geophysical Research Letters. DOI:10.1029/2019GL082410.

Clark, A. J., and Coauthors, 2012: An Overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

Clark, A.J. and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.,* **99**, https://doi.org/10.1175/BAMS-D-16-0309.1

Fierro, A. O., Zhao G., Liu S., Wang Y., J. Gao, K. Calhoun, C. L. Ziegler, E. R. Mansell and D. R. MacGorman, 2018: Assimilation of total lightning with GSI and NEWS3DVAR to improve short-term forecasts of high impact weather events at cloud resolving scales. *Accepted to the JCSDA Quarterly Newsletter, Winter 2017.*

Fierro, A. O., J. Gao, C. Ziegler, K. Calhoun, E.R. Mansell and D. R. MacGorman 2016: Assimilation of pseudo GLM flash extent data in the variational framework at convection allowing scales: Proof-of-concept and evaluation. *Mon. Wea. Rev.,* **144**, 4373– 4393.

Gagne, D. J., A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-Based Probabilistic Hail Forecasting with Machine Learning Applied to Convection-Allowing Ensembles. *Wea. Forecasting*, **32**, 1819-1840.

Gallo, B.T., and Coauthors, 2017: Breaking New Ground in Severe Weather Prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting,* **32**, 1541–1568, https://doi.org/10.1175/WAF-D-16-0178.1

Gallo, B. T., A. J. Clark, B.T. Smith, R.L. Thompson, I. Jirak, and S.R. Dembek, 2018: Blended Probabilistic Tornado Forecasts: Combining Climatological Frequencies with NSSL–WRF Ensemble Forecasts. *Wea. Forecasting,* **33**, 443–460, https://doi.org/10.1175/WAF-D-17-0132.1.

Gao, J., T. M. Smith, D. J. Stensrud, C. Fu, K. Calhoun, K. L. Monross, J. Brogden, V. Lakshmanan, Y. Wang, K. W. Thomas, K. Brewster, and M. Xue, 2013: A Real-Time Weather-Adaptive 3DVAR Analysis System for Severe Weather Detections and Warnings. *Wea. Forecasting*, **28**, 727-745.

Harris, L. M., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev*., **141**, 283-306.

Harris, L.M., S.-J. Lin, and C.-Y. Tu, 2016: High-Resolution Climate Simulations Using GFDL HiRAM with a Stretched Global Grid. *Journal of Climate*, **29**, 4293-4314.

Hong, S.Y., Noh, Y. and Dudhia, J., 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.,* **34**, 2318-2341.

Jirak, I. L., C. J. Melick, A. R. Dean, S. J. Weiss, and J. Correia, Jr., 2012: Investigation of an automated temporal disaggregation technique for convective outlooks during the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. Preprints, *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 10.2.

Jirak, I. L., C. J. Melick, and S. J. Weiss, 2014: Combining probabilistic ensemble information from the environment with simulated storm attributes to generate calibrated probabilities of severe weather hazards. Preprints, *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 2.5.

Jones, T. A., P. Skinner, N. Yussouf, K. Knopfmeier, A. Reinhard, X. Wang, K. Bedka, W. Smith, Jr., and R. Palikonda, 2020: Assimilation of Goes-16 Radiances and Retrievals into the Warn-on-Forecast System. *Mon. Wea. Rev*., **148**, 1829-1859.

Loken, E. D., A. J. Clark, and C. D. Karstens, 2020: Generating Probabilistic Next-Day Severe Weather Forecasts from Convection-Allowing Ensembles Using Random Forests. *Wea. Forecasting* (In review).

Rothfusz, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. D. Karsten, G. J. Stumpf, and t. M. Smith, 2018: FACETs: A Proposed Next-Generation Paradigm for High-Impact Weather Forecasting. *Bull. Amer. Meteor. Soc.,* **99**, 2025-2043.

Stensrud, D. J., and Coauthors, 2009: Convective-Scale Warn-on-Forecast System. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

Thompson, G. and T. Eidhammer, 2014: A Study of Aerosol Impacts on Clouds and Precipitation Development in a Large Winter Cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1.

Wang, Y., J. Gao, P. S. Skinner, K. Knopfmeier, T. Jones, G. Creager, P. L. Heinselman, and L. J. Wicker, 2019: Test of a Weather-Adaptive Dual-Resolution Hybrid Warn-on-Forecast Analysis and Forecast System for Severe Severe Weather Events. *Wea. Forecasting*, **34**, 1807-1827.

Weiss, S. J., A. J. Clark, I. L. Jirak, C. J. Melick, C. Siewert, R. A. Sobash, P. T. Marsh, A. R. Dean, J. S. Kain, M. C. Coniglio, M. Xue, F. Kong, K. W. Thomas, J. Du, D. R. Novak, F. Barthold, M. J. Bodner, J. J. Levit, C. B. Entwistle, R. S. Schneider, and T. L. Jensen, 2010: An Overview of the 2010 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. *25th Conf. on Severe Local Storms*, Amer. Meteor. Soc., 7B.1.

Yussouf, N., and K. H. Knopfmeier, 2019: Application of the Warn-on-Forecast System for flash-flood-producing heavy convective rainfall events. *QJRMS*, **145**, 2385-2403.

Zhou, L., S.-J. Lin, J.-H. Chen, L. Harris, X. Chen, and S. L Rees, 2019: Toward Convective-Scale Prediction within the Next Generation Global Prediction System. *Bull. Amer. Meteor. Soc*. DOI:10.1175/BAMS-D-17-0246.1.